Quest for consistency, symmetry and simplicity - The legacy of Albert Tarantola

Mosegaard, Klaus

Published in:
Geophysics

Link to article, DOI:
10.1190/GEO2010-0328.1

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

GEOPHYSICS. VOL. 76, NO. 5 (SEPTEMBER-OCTOBER 2011); P. W51–W61, 10 FIGS.
10.1190/GEO2010-0328.1

Quest for consistency, symmetry, and simplicity — The legacy of Albert Tarantola

Klaus Mosegaard[1]

ABSTRACT

On 6 December 2009, the distinguished Spanish-French physicist and geoscientist, Albert Tarantola, passed away at the age of 60. Born in Barcelona in 1949, he went to Paris where he lived most of his life, and worked as a professor at Institut de Physique du Globe de Paris. His extensive scientific production and remarkable achievements in inverse problem theory and geophysical data analysis established him as one of the most influential mathematical geoscientists of our time. He became the father of probabilistic inverse theory, a theory that he passionately defended against severe criticism and managed to propagate to a major part of the geophysical community. Another major achievement was his contributions to the theory of seismic waveform inversion — a work that right now is unfolding its potential in large-scale computations. Albert's contributions were not limited to geoscience. He started his career in astrophysics, and later in his life he wrote several papers and books on physics and probability, including new formulations of fluid dynamics, elasticity theory, global positioning, and scientific inference. Albert possessed a unique combination of exceptional skills and remarkable mental energy. He was a veritable powerhouse with an unusual work ethic, and his passion for science will remain an ideal for all those who worked with him.

INTRODUCTION

Albert Tarantola died 6 December 2009, a couple of months after a sudden and totally unexpected stroke while on a visit to Chile, where he lectured at University of Santiago. Albert was born in Barcelona in 1949, but he spent most of his life in Paris, where he unfolded his unusual talent for science and mathematics. After completing university studies in Barcelona, he went to Paris where he in 1976 wrote a Ph.D. thesis on theoretical astrophysics at Université de Paris 6. The thesis dealt with the evolution of clusters of galaxies, using the general theory of relativity, and was driven by a desire to gain a deep understanding of the structure of the universe. However, while working on his thesis, Albert realized that science is still far from achieving this goal. As a result, he turned his interests toward inverse problems, a topic of importance and profound philosophical implications, not only to geophysics, but also to the physical sciences in general.

Albert was appointed professor at Institut de Physique du Globe de Paris in 1983, and in 1991 he was promoted to professor in the category *classe exceptionnelle*. Over the years, he received several awards for his work on inverse problems based on a probabilistic foundation. He formed a research group — The Geophysical Tomography Group — which, during the years 1985–2000, made pioneering work on nonlinear inversion of seismic waveforms. In this fertile period, Albert and his group made many novel contributions of both theoretical and practical importance, and he published the first edition of his widely known book *Inverse Problem Theory* (first edition: Elsevier, 1987, second edition: SIAM, 2005), a book which can be found on the bookshelves of many geoscientists all over the world. The book is recommended reading for any scientist seeking details from the "engine room" of geophysical inversion in general, and seismic inversion in particular.

PURE PROBABILISM

To appreciate Albert Tarantola's impact on computational geoscience, we must understand the underlying motives and interests behind his extensive scientific production. In all the problems he worked on, he strived toward a complete and invariant physical description, based on a simple mathematical formalism without unnecessary concepts. This consequent reductionism was guided by a strong eye for symmetries, and supplemented by a systematic avoidance of inconsistencies.

Albert's most important writings were dedicated to inverse problem theory, the system of ideas and methods used to infer

Manuscript received by the Editor 11 October 2010; revised manuscript received 27 January 2011; published online 7 November 2011.
[1]Technical University of Denmark, Department of Informatics and Mathematical Modeling. Lyngby, Denmark. E-mail:kmos@imm.dtu.dk.

information about a physical system, e.g., the Earth's interior, from observational data and any other prior information. The physical system is described by a set of model parameters $\mathbf{m} = (m_1, m_2, \ldots, m_M)$ or spatial functions $\mathbf{m}(\mathbf{x}) = (m_1(\mathbf{x}), m_2(\mathbf{x}), \ldots, m_M(\mathbf{x}))$, usually denoted *the model*, and a set of observations, *the data*, $\mathbf{d} = (d_1, d_2, \ldots, d_N)$ which are theoretically related to the model via a set of equations

$$\mathbf{d} = g(\mathbf{m}). \tag{1}$$

In addition to this, there is usually some information available on data uncertainties, and a priori constraints on the model. The latter may or may not be formulated in probabilistic terms. Throughout his career, Albert was mostly interested in the seismic problem for which $g(\mathbf{m})$ is usually a highly nonlinear function, but his theoretical work was very general, covering all types of inverse problems.

From the very outset, Albert was dissatisfied about inverse theory as it was formulated by the end of the 1970s. In his opinion it was an ungraceful mixture of highly diverse and often scientifically unjustified concepts: arbitrary regularizations, arbitrary parameterizations, disregarded uncertainties, and not least prior information that was generally discounted.

By the end of the 1970s, geophysical inverse theory was dominated by two main methods aimed at linear or linearized problems, namely the *regularization methods* (Levenberg, 1944; Tikhonov, 1963; Marquardt, 1970), and the method of Backus and Gilbert (Backus and Gilbert, 1967, 1970). Each of these methods had its own way of dealing with nonuniqueness, that is, multiple (often infinitely many) solutions to an inverse problem. If a linear inverse problem

$$\mathbf{d} = \mathbf{G}\mathbf{m}, \tag{2}$$

where data $\mathbf{d}$ and unknown model parameters $\mathbf{m}$ are related through a linear operator $\mathbf{G}$, was underdetermined, the inverse of $\mathbf{G}$ did not exist, but a unique "acceptable" solution could be generated through regularization. In its simplest form, the regularized solution was given by the expression (Tikhonov, 1963):

$$\mathbf{m}_T = (\mathbf{G}^t\mathbf{G} + \lambda\mathbf{I})^{-1}\mathbf{G}^t\mathbf{d} \tag{3}$$

where λ was the so-called *regularization parameter*, an adjustable parameter found after some experimentation. The criterion for choosing a value of λ was usually that the resulting solution $\mathbf{m}_T$ looked "reasonable." The regularization method, which is still very popular today, was considered unscientific by Albert. The choice of regularization parameter was crucial for the method, but, in his opinion, completely arbitrary. Clearly, some sort of prior information about the solution was needed to choose λ, but it was unclear how λ was related to that information. Added to this, nonuniqueness of the solution due to uncertainty of the data was not dealt with by the method and had to be analyzed separately.

The Backus-Gilbert method dealt with nonuniqueness in an entirely different way. Instead of choosing one of many possible solutions, as in Tikhonov's method, Backus and Gilbert avoided a priori constraints and instead devised a method for calculating unique localized averages. A Backus-Gilbert solution $\mathbf{m}_{\mathrm{BG}}$ to (2) was a linear function of the data $\mathbf{d}$, designed to jointly minimize the variance of the model uncertainty, and maximize the resolution in the sense that if the data $\mathbf{d}$ were generated by a delta function $\mathbf{m}_\delta$, then $\mathbf{m}_{\mathrm{BG}}$ would be as close as possible to $\mathbf{m}_\delta$ in some predefined sense.

Albert found that this method had three drawbacks: (1) it was not obvious how to extend the method to general, nonlinear inverse problems, (2) nonuniqueness of the solution due to data uncertainties were not dealt with, and (3) it avoided the use of prior information. Backus and Gilbert defended a non-Bayesian standpoint, considering prior information as an unwanted "subjective" bias. Their viewpoint was supported by the fact that, in practice, most Bayesian computations relied on arbitrary priors, usually chosen for purposes of mathematical convenience and simplicity. Albert's vision was to use priors that were rooted in earlier observations or, as recommended by Jeffreys (1939), in symmetry arguments, rather than in pure subjective belief. In this way he rejected the widespread use of arbitrary priors. To realize his vision of probabilistic inverse theory, Albert felt that there was a need for a method that could combine information from several independent sources: new data, old data (geophysical or other kinds), geological experience, symmetry considerations, etc.

This position became the starting point for his and Bernard Valette's work on a new formulation of inverse theory, first published in the seminal paper "Inverse problems = quest for information" (Tarantola and Valette, 1982a). The paper represented a radical departure from established theory. It was axiomatic, and it insisted on a purely probabilistic formulation where physical parameters were fully represented, not by numbers, but by probability distributions. Moreover, the same probability distributions were assumed to be a complete description of the state of information we have on the corresponding parameters. Solving the inverse problem could now be formulated as an act of combining different and independent states of information on the model parameters, some of which came from observations, others from prior information.

Based on a set of simple and natural axioms, Tarantola and Valette derived a mathematical expression for the combination of states of information on a set of parameters. This expression could be used to combine prior (data-independent) information on the parameters, and information contained in the data. Assuming that information from two independent sources of information on a set of parameters $\mathbf{x}$ are described by the probability densities $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, respectively, Tarantola and Valette found that the two pieces of independent information could be combined into a new probability density $f(\mathbf{x})$ through the simple equation

$$f(\mathbf{x}) = \frac{f_1(\mathbf{x}) f_2(\mathbf{x})}{\mu(\mathbf{x})} \tag{4}$$

where the (everywhere positive) probability density $\mu(\mathbf{x})$ represented null information. The mathematical object $\mu(\mathbf{x})$ was new and controversial in inverse problem theory. It became the focus of a heated debate between Albert and proponents of the "frequentist" school of thought that dominated the field at that time.

Albert's critics claimed that $\mu(\mathbf{x})$ could not be an adequate description of null information. In their opinion, the simple, nonprobabilistic statement: "The solution resides in a certain subvolume $\mathcal{A}$" carried less information than a distribution that equaled the "noninformative" probability density $\mu(\mathbf{x})$ in the subvolume $\mathcal{A}$, and was zero outside. $\mu(\mathbf{x})$ would, in addition, allocate equal probabilities to equal volumes in $\mathcal{A}$ (see, e.g., Backus, 1988).

Albert's notion of information was based entirely on Shannon's probabilistic definition (Shannon, 1948), and he found that his critics' mixing of probabilistic concepts of information with nonprobabilistic concepts was meaningless. He resisted their arguments for several years, but eventually he complied to some extent with the criticism and changed the name of $\mu(\mathbf{x})$ to *"the homogeneous probability density."* This term was first used in Mosegaard and Tarantola (2002), a precursor to the second edition of Albert's book on inverse problem theory (Tarantola, 2005). In retrospect, the controversy about $\mu(\mathbf{x})$ was perhaps one of the few cases where he gave up prematurely. A correct resolution of this conflict would have required that the two parties at least had agreed on a definition of the notion of "information," but such an agreement never materialized.

Albert had a good reason for defending his introduction of $\mu(\mathbf{x})$: it was crucial for the use of equation 4 to solving inverse problems. When information, given by a probability density $f(\mathbf{x})$, was augmented with null information, the result should be $f(\mathbf{x})$, perhaps except for a constant. In other words, $f_2(\mathbf{x}) = \mu(\mathbf{x})$ should lead to $f(\mathbf{x}) = f_1(\mathbf{x})$. Only if $\mu(\mathbf{x})$ was always constant could we do without it in the denominator, but Albert pointed to a number of cases where the assumption of a constant $\mu(\mathbf{x})$ could not be maintained. The most notable examples are the *Jeffreys parameters*, parameters ν which are inherently positive and whose reciprocal parameters $1/\nu$ are also commonly employed. Well-known examples are electrical conductivity (the reciprocal of electrical resistivity) and acoustic velocity (the reciprocal of acoustic slowness).

Jeffreys parameters ν could not have a constant $\mu(\nu)$, because the probability that ν lies between ν_1 and ν_2 should equal the probability that its reciprocal $\chi = 1/\nu$ lies between $1/\nu_1$ and $1/\nu_2$. Jeffreys (1939) had advocated the use of $\mu(\nu) = \text{const}/\nu$ and $\mu(\chi) = \text{const}/\chi$ as symmetric definitions of homogeneous probability density for ν and χ to satisfy this requirement.

To Albert, the requirement (symmetry) that homogeneous probability densities for mutually reciprocal parameters had the same mathematical form was important. He started his career working with general relativity in astrophysics, a theory that grew out of Einstein's vision of formulating physical laws that were invariant under a change of reference frame. With equation 4, and by insisting on nonconstant homogeneous probability densities, he had reached a similar goal in inverse problem theory.

UNCERTAIN THEORIES

In Tarantola and Valette's most general formulation of inverse problem theory, based on equation 4, they retained the possibility that the relation $\mathbf{d} = g(\mathbf{m})$ was not known exactly. This was an entirely new concept in inverse theory, and one that is still quite advanced. Let us assume that the inverse problem is characterized by the following three elements: (1) a so-called prior probability distribution $\rho_m(\mathbf{m})$ carrying the information we have on $\mathbf{m}$ before any new data are analyzed, (2) uncertain observations, described by a probability distribution $\rho_d(\mathbf{d})$, from which our observed data set $\mathbf{d}_{\text{obs}}$ is a realization, and finally (3) a possibly uncertain relation $\mathbf{d} \approx g(\mathbf{m})$ described by a distribution $\theta(\mathbf{d}, \mathbf{m})$ over the joint data/model space. If we assume that the prior information on the model parameters is data-independent, the joint prior satisfies $\rho(\mathbf{d}, \mathbf{m}) = \rho_d(\mathbf{d})\rho_m(\mathbf{m})$, and we can now use equation 4 to combine all the information we have on the inverse problem, which is contained in $\rho(\mathbf{d}, \mathbf{m})$ and $\theta(\mathbf{d}, \mathbf{m})$. The result is the so-called *posterior probability density* (see Figure 1):

$$\sigma(\mathbf{d}, \mathbf{m}) = \frac{\rho(\mathbf{d}, \mathbf{m})\theta(\mathbf{d}, \mathbf{m})}{\mu(\mathbf{d}, \mathbf{m})}. \tag{5}$$

In many applications this leads to

$$\sigma_m(\mathbf{m}) = k\rho_m(\mathbf{m})L(\mathbf{m}) \tag{6}$$

where

$$\sigma_m(\mathbf{m}) = \int_D \sigma(\mathbf{d}, \mathbf{m}) d\mathbf{d} \tag{7}$$

is the marginal posterior probability distribution over the model space, and where the *likelihood function* $L(\mathbf{m})$, measuring the model's fit to the observed data, is given by

$$L(\mathbf{m}) = \int_D \frac{\rho_d(\mathbf{d})\theta(\mathbf{d}|\mathbf{m})}{\mu_d(\mathbf{d})} d\mathbf{d}. \tag{8}$$

Here we have assumed that $\theta(\mathbf{d}, \mathbf{m}) = \theta(\mathbf{d}|\mathbf{m})\mu_m(\mathbf{m})$, meaning that $\theta(\mathbf{d}, \mathbf{m})$ expresses the correlation between $\mathbf{d}$ and $\mathbf{m}$, but no information about $\mathbf{m}$. Another assumption behind equation 6 is that the homogeneous probability density $\mu(\mathbf{d}, \mathbf{m})$ in the joint model/data space can be simplified to $\mu(\mathbf{d}, \mathbf{m}) = \mu_d(\mathbf{d})\mu_m(\mathbf{m})$.

The concept of "uncertain theories" is important in applied physics in general, and in geoscience in particular. For instance, in seismic inversion, approximate modeling schemes are widely used to cut down on the computational work needed to solve large-scale problems. It is fully recognized that these approximations influence the result, but unfortunately it is not clear how one could obtain reasonable modelization uncertainties, and how realizations of the uncertainties could be obtained. Today, it is still an open question whether progress can been made in specifying uncertainties in, e.g., seismic modeling, but it is Albert's groundbreaking theoretical work that forms the basis for this research area and paves the way for future solutions.

SIDESTEPPING BAYES

In the special case where we have an exact relation $\mathbf{d} = g(\mathbf{m})$ between model parameters and data, it is tempting to put $\theta(\mathbf{d}|\mathbf{m}) = \delta(\mathbf{d} - g(\mathbf{m}))$, where δ is a delta function, and to set

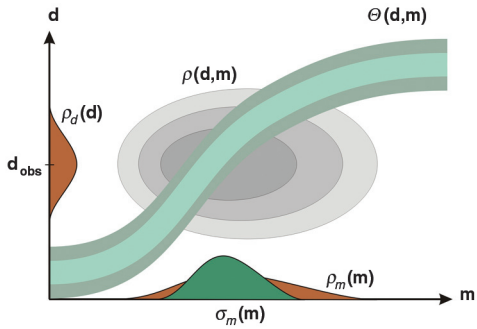

Figure 1. Tarantola and Valette (1982a) saw inversion as a combination of independent pieces of information, describing the prior information on data and unknown parameters, $\rho(\mathbf{d}, \mathbf{m})$, and the information carried by a theoretical relation between data and parameters, $\theta(\mathbf{d}, \mathbf{m})$.

$\mu_d(g(\mathbf{m}))$ equal to a constant, after which the likelihood function (8) simplifies to

$$L(\mathbf{m}) = \rho_d(g(\mathbf{m})). \tag{9}$$

In this case, equation 6 becomes

$$\sigma_m(\mathbf{m}) = k\rho_m(\mathbf{m})\rho_d(g(\mathbf{m})). \tag{10}$$

If we change the notation slightly, introducing conditioning on $\mathbf{d}_{\text{obs}}$ explicitly through the identifications $f(\mathbf{m}|\mathbf{d}_{\text{obs}}) \equiv \sigma_m(\mathbf{m})$, $f(\mathbf{d}_{\text{obs}}|\mathbf{m}) \equiv \rho_d(g(\mathbf{m}))$, $f(\mathbf{m}) \equiv \rho(\mathbf{m})$, and $f(\mathbf{d}_{\text{obs}}) \equiv 1/k$ we obtain *Bayes' Theorem*:

$$f(\mathbf{m}|\mathbf{d}_{\text{obs}}) = \frac{f(\mathbf{m})f(\mathbf{d}_{\text{obs}}|\mathbf{m})}{f(\mathbf{d}_{\text{obs}})}. \tag{11}$$

relating a posterior probability density $f(\mathbf{m}|\mathbf{d}_{\text{obs}})$ to a prior probability density $f(\mathbf{m})$, and a likelihood function $f(\mathbf{d}_{\text{obs}}|\mathbf{m})$. As $\mathbf{d}_{\text{obs}}$ is a constant vector, $f(\mathbf{d}_{\text{obs}})$ is just a constant.

For many years, Bayes' theorem has been widely used as the main tool for probabilistic inversion, but Albert's analysis revealed an unpleasant inconsistency in this theorem. The inconsistency came from the operation of putting $\theta(\mathbf{d}|\mathbf{m}) = \delta(\mathbf{d} - g(\mathbf{m}))$, which corresponds to forming a conditional probability density. This can be done in several ways, as illustrated in Figure 2. Each possibility is characterized by its own way of obtaining $\delta(\mathbf{d} - g(\mathbf{m}))$ as a limit of "narrowing" a distribution that is constant in a corridor surrounding the curve $\mathbf{d} = g(\mathbf{m})$. The upper part of the figure shows the result of letting the upper and lower boundaries of the corridor approach vertically, whereas the lower part shows a limit where the boundaries converge by keeping a constant width of the corridor. The problem here is that the two limits give different results for $\sigma(\mathbf{m})$.

The standard procedure of Bayes' theorem, where we simply put $\mathbf{d} = g(\mathbf{m})$,

$$f(\mathbf{x}|\mathbf{y} = \mathbf{y}_0) = \frac{f(\mathbf{x}, \mathbf{y}_0)}{f(\mathbf{y}_0)} \tag{12}$$

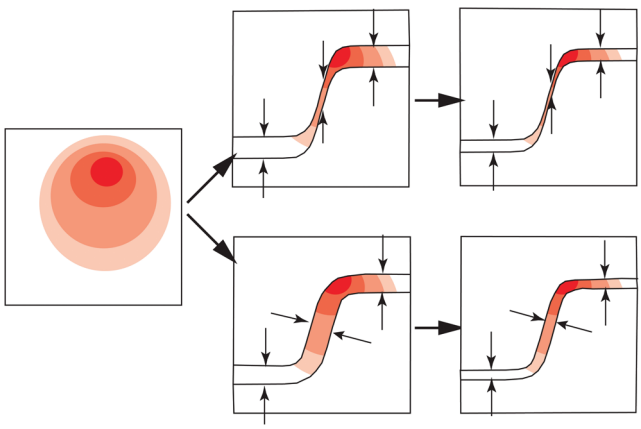

Figure 2. The definition of a conditional probability density as the limit of the probability mass along a corridor whose width tends to zero. The fact that such a limit can be obtained in many different ways makes the notion of a conditional probability density ambiguous.

corresponds to a special limit operation where the width of the corridor decreases "along cordinate axes." A simple example shows how this traditional definition of conditional probability density may lead to a self-contradiction:

Consider a situation where we are told that a meteorite will hit the Earth's surface, and that the probability distribution (probability per unit area) for the impact point is constant everywhere on the surface. If we recall that the probability density $f(\theta, \phi)$ as a function of latitude θ and longitude ϕ is defined such that the probability of having an impact in an area $d\theta d\phi$ around point (θ, ϕ) is

$$f(\theta, \phi)d\theta d\phi, \tag{13}$$

it is easily shown that $f(\theta, \phi) = \cos\theta/4\pi$. If we subsequently learned that new calculations had revealed that the meteorite would, in fact, hit the Earth at, say, $\phi = \phi_0$, it would be tempting to conclude (following the thinking behind Bayes' theorem) that the probability density we now have for θ would be

$$f(\theta|\phi = \phi_0) = \frac{f(\theta, \phi_0)}{f(\phi_0)} = \frac{\cos\theta}{2} \tag{14}$$

where

$$f(\phi_0) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(\theta, \phi_0)\mathrm{d}\theta. \tag{15}$$

The probability density 14 is nonconstant, and this is obviously a false conclusion: the probability that the meteorite will hit points on the $\phi = \phi_0$ meridian is obviously constant over the meridian. We have run into the so-called *Borel paradox*.

The reason for the self-contradictory result is the hidden limiting procedure in Bayes' theorem, where the corridor around the $\phi = \phi_0$ meridian is narrowed while keeping its angular width constant (with boundaries of constant longitudes). Had we instead kept the physical width (in km) of the corridor constant in the limiting process, we would have obtained a constant distribution on the meridian, and no self-contradiction would have occurred.

From this, Albert concluded that the notion of conditional probability density is inconsistent, and that the correct solution of an inverse problem requires specification of a nonsingular $\theta(\mathbf{d}, \mathbf{m})$.

PRAGMATISM

The work of Tarantola and Valette (1982a) was followed by a companion paper (Tarantola and Valette, 1982b) where the theory was applied to linear and weakly nonlinear problems with Gaussian noise and Gaussian prior. This paper was dedicated to the method of least-squares, which was, in fact, later criticized strongly by Albert. We shall come back to this criticism later, but first let us take a look at the main results of the paper. For purely linear, Gaussian problems, Tarantola and Valette arrived at formulas similar to those of the so-called method of *stochastic inversion* (Franklin, 1970). However, they also derived iterative formulas for solving weakly nonlinear problems, a method later known as the *Tarantola-Valette method*. The formulas expressed how an initial model $\mathbf{m}_1$ numerically evolves into a solution close to the one with maximum value of the posterior probability density. Each iteration is typically based on the equation

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \varepsilon_k(\mathbf{C}_M\mathbf{F}_k^t\mathbf{C}_D^{-1}(\mathbf{d}_k - \mathbf{d}_{\text{obs}}) + (\mathbf{m}_k - \mathbf{m}_{\text{prior}})) \tag{16}$$

where $\mathbf{d}_{\text{obs}}$ is the observed data, $\mathbf{m}_{\text{prior}}$ is the prior mean model (sometimes called the "prior" model), $\mathbf{m}_k$ is the kth approximation to the model, $\mathbf{d}_k$ is the kth computed data, $\mathbf{C}_D$ and $\mathbf{C}_M$ are covariance matrices for data and prior, respectively, and $\mathbf{F}$ is the matrix of derivatives of data with respect to the model parameters. The covariance matrix for the posterior probability density is evaluated from

$$\tilde{\mathbf{C}}_M \approx (\mathbf{F}_\infty^t\mathbf{C}_D^{-1}\mathbf{F}_\infty + \mathbf{C}_M^{-1})^{-1}, \tag{17}$$

where $\mathbf{F}_\infty$ refers to the value of the matrix of partial derivatives at the convergence point.

Equations 16 and 17 were derived under a simplifying assumption of constant $\mu_d(\mathbf{d})$ and $\mu_m(\mathbf{m})$. The method of Tarantola and Valette, expressed through equation 16, looks like a straightforward application of a steepest ascent algorithm to maximize the posterior probability, but in cases where μ_d and μ_m are not constant (e.g., in seismic problems) their theory calls for, not a maximation of the posterior $\sigma(\mathbf{m})$, but rather a maximation of the ratio

$$\frac{\sigma_m(\mathbf{m})}{\mu_m(\mathbf{m})}. \tag{18}$$

Albert had discovered that maximizing $\sigma(\mathbf{m})$, as usually recommended in the literature, may lead to different results in different model parameterizations. Consider an acoustic (or simplified seismic) inverse problem where the unknown subsurface structure $\mathbf{v}$ is parameterized as wave velocities in a set of grid points. Assume that we, for given prior information $\rho_v(\mathbf{v})$, invert the data by maximizing $\sigma_v(\mathbf{v})$ and obtain a solution $\hat{\mathbf{v}}$. If, for some reason, we now change our parameterization from velocities v_i to slownesses $s_i = 1/v_i$, transform the prior from ρ_v to ρ_s, and seek the solution $\hat{\mathbf{s}}$ in terms of slowness, maximizing $\sigma(\mathbf{s})$ would provide a different answer (inconsistent with the first solution $\sigma_v(\mathbf{v})$): for each slowness parameter $\hat{s}^i$, the relation $\hat{s}^i = 1/\hat{v}^i$ would, in general, not be satisfied. If we instead realize that velocities and slownesses are Jeffreys parameters and introduce the homogeneous probability densities

$$\mu_v(\mathbf{v}) = \prod_{k=1}^{M}\frac{1}{v_k} \quad \mu_s(\mathbf{s}) = \prod_{k=1}^{M}\frac{1}{s_k}, \tag{19}$$

the inconsistency disappears. The same thing happens if we, instead of working with the original parameters, work with logarithmic parameters $v_k' = \log(v_k/v_0)$ and $s_k' = \log(s_k/s_0)$, where v_0 is a reference velocity and s_0 is a reference slowness. It is easy to demonstrate that such logarithmic parameters have constant homogeneous probability densities. To Albert, the above considerations confirmed the need for accepting $\mu_v(\mathbf{v})$ and $\mu_s(\mathbf{s})$ as homogeneous probability densities for $\mathbf{v}$ and $\mathbf{s}$, respectively.

It is, however, interesting that, among practitioners adhering to Albert's probabilistic inversion, μ is rarely assumed nonconstant for Jeffreys parameters. Despite the inconsistency, there seems to be a psychological barrier preventing numerical modelers from weighing small values of, e.g., velocity and slowness higher than large values.

A MOST DIFFICULT PROBLEM

Albert was very consequent in his striving toward creating a consistent theory, but at critical moments in his career he became very pragmatic. After his work on the two 1982 inversion papers, his vision was to make serious progress in one of the most important and ambitious fields of applied geophysics: inversion of seismic reflection data. This field was (and is) characterized by huge computational demands, so Albert realized that simplifications were unavoidable.

In a series of seminal papers (Tarantola, 1984, 1986, and 1988) he carefully described how iterative nonlinear waveform inversion of general reflection data from a 3D-subsurface could be carried out, using the formalism outlined in Tarantola and Valette (1982b). Albert was influenced by Jon Claerbout, Patrick Lailly, and others (Claerbout, 1971; Lailly, 1983), but he was the first to undertake the ambitious project of developing a general theory of waveform inversion where the goal was to carefully reproduce all details of the observed field (Lailly, 2010). Starting with the purely acoustic case (Tarantola, 1984), he worked his way through the problem, removing one approximation after the other, crowning his work with the paper "Theoretical Background for the Inversion of Seismic Waveforms, Including Elasticity and Attenuation" (Tarantola, 1988).

In each paper in this series, a steepest descent algorithm minimizing the misfit function was developed, but the calculations were supplemented with a new and important insight: each iteration consisted of a propagation of the actual source in the current medium, a propagation of the residuals — acting as if they were sources — backward in time, and a correlation of the resulting two wave fields. Remarkably, this interpretation of each step of the algorithm closely resembled the well-known imaging principle which was the backbone of Jon Claerbout's theory of migration of seismic reflection data (Claerbout, 1971). It gave a unique physical insight into the problem, a very characteristic element in Albert's approach to science.

Albert's method of using propagated waveform residuals backward in time as a means of determining the gradient of the misfit function is an example of what has later become known as the "adjoint method" (Talagrand and Courtier, 1987) a very important concept in modern seismic data analysis.

A crucial property of Albert's method was that only two wavefields should be computed for each shot. This property significantly cut down the computational workload required by the inversion, and was instrumental in helping the subsequent numerical experiments materialize. The computer resources of that time were barely sufficient for such experiments, but Albert and his research group — The Geophysical Tomography Group — had some success in the late 1980s with the first, careful full-waveform inversions (see, e.g., Crase et al., 1992). A later example of inversions by Albert's group (Charara et al., 1996, 2000; Barnes et al., 1998; Tarantola, 2005) is shown in Figures 3–7b reproduced from Tarantola (2005). Three-component seismometers, located between 2 and 4 kilometers depth in a borehole, record seismic waves generated at the surface (Figure 3). The source is offset 2 km from the borehole (see a sketch of the measurement geometry in Figure 4). The a priori model $\mathbf{m}_{\text{prior}}$ is horizontally layered. The subsurface model obtained from a viscoelastic, full-waveform inversion is seen in Figures 5a–6b. It is seen from the data residuals in Figure 7b that most of the coherent energy in the data is modeled in the inversion

process, although some artifacts (e.g., imprints of "raypaths" in the *P*-velocity model of Figure 5a) are present in the final model.

During the 1980s, Albert pointed out that a vital part of seismic inversion problems cannot be solved by iterative algorithms like 16, unless you have a starting model that is close to the true solution. The dependence of the data on the long wavelengths of the velocity field is so nonlinear that steepest descent methods and other local optimization strategies are likely to fail.

EMBRACING RANDOMNESS

In fact, the results displayed in Figures 4, 5, 6, and 7 would not have been obtained without an initial Monte Carlo search for a reasonable velocity model fitting most of the arrival times of major events (see Barnes et al., 1998). Once this was done, a preconditioned steepest descent algorithm was able to provide an acceptable model in a few tens of iterations (see Charara et al., 1996, 2000).

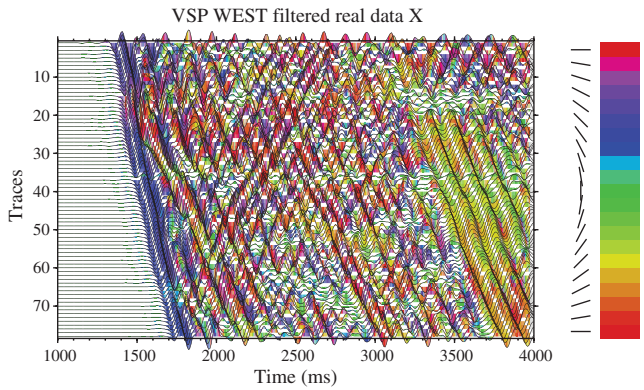

Figure 3. Horizontal and vertical component data were used in the full waveform inversion of Charara et al. (1996, 1998, 2000). The figure shows the horizontal component only. The instantaneous polarization of the wave is shown by the color code.

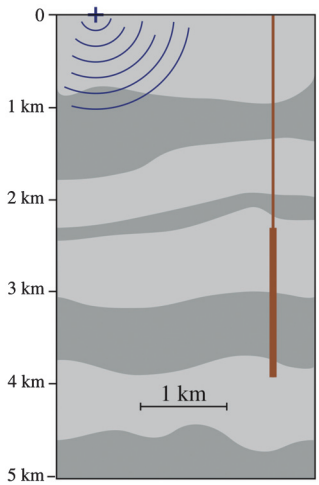

Figure 4. A sketch of the measurement geometry in the full-waveform inversion example of Charara et al. (1996, 2000), and Barnes et al., 1998.

In the beginning, Albert saw Monte Carlo methods as a means of finding optimal solutions to the "background velocity problem," the problem of finding the best-fitting long-wavelength velocity field. For this reason he was strongly attracted to simulated annealing, genetic algorithms, and similar stochastic optimization methods.

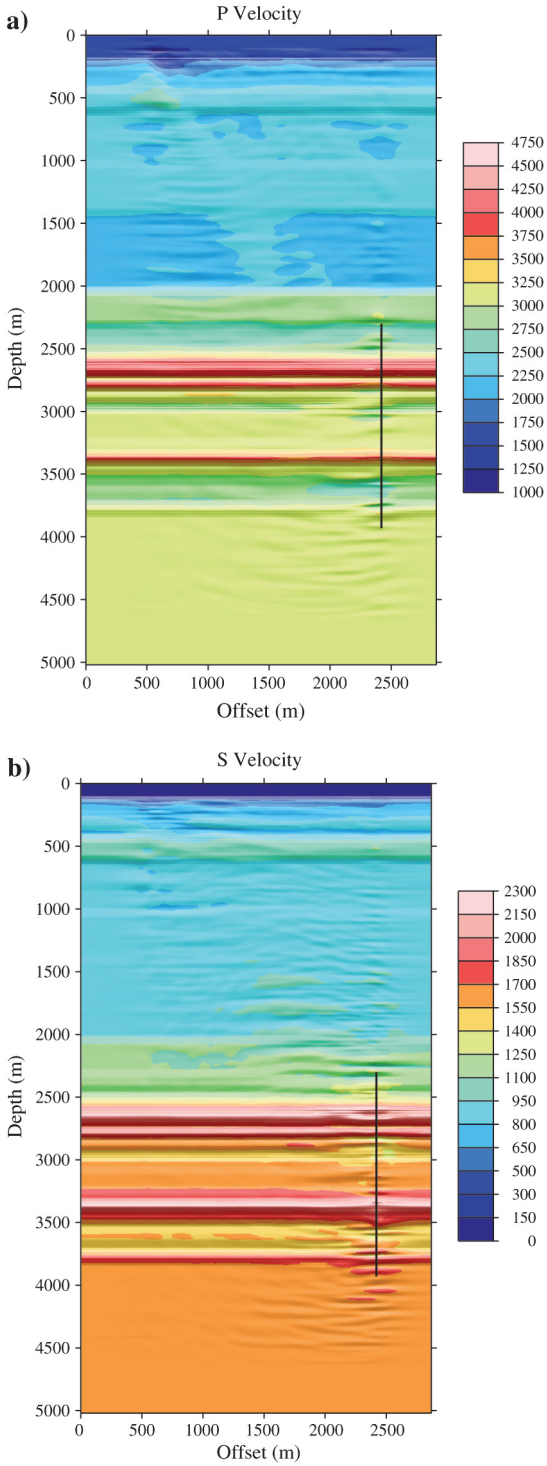

Figure 5. P-velocity model (top) and S-velocity model (bottom) obtained from waveform inversion of the data (and the corresponding vertical component data). The velocity unit is m/s.

In the late 1980s, the use of these methods and the wider class of Markov Chain Monte Carlo (MCMC) methods in large-scale computations emerged as a real possibility (see, e.g., Rothman, 1986; Vestergaard and Mosegaard, 1991), and Albert was fascinated by their potential. Furthermore, he realized that Monte Carlo methods could actually help bridge the gap between his and Valette's original — and very general — theory, and more practical nonlinear inversion. The reason is that MCMC methods allowed probability densities to be sampled, thereby generating, e.g., subsurface models with a density in model space proportional to the posterior probability density. Albert saw that a collection of such sample models would in themselves constitute a far better description of a strongly non-Gaussian posterior probability than just reporting a mean model and a covariance matrix.

In the following years, he worked hard on this idea. A first paper, demonstrating the use of MCMC methods for analysis of nonlinear problems, was published by Koren et al. (1991). This paper described a Monte Carlo solution to the background velocity problem. For each velocity model, synthetic travel times were calculated, and a waveform misfit in corridors along synthetic traveltime curves was computed. Each background velocity field generated by the MCMC algorithm was used to produce a migrated stack which was subsequently displayed. Using this procedure, Koren, et al., realized, for the first time, the vision of generating a sample of posterior models for a geophysical problem of a (for that time) reasonable scale. At the same time, it was a first demonstration of the movie strategy, where the many sample models were presented to the spectator in the form of a sequential display of images (see Figure 8) In fact, the movie strategy was a first example of a method which was later described in computer visualization science (e.g., Zuk, and Carpendale, 2006).

The sampling strategy for nonlinear inversion was described in detail in a paper by Mosegaard and Tarantola (1995). It was essentially similar to Bayesian computation (see, e.g., Bernardo and Smith, 1994), except that it was based on the covariant theory of Tarantola and Valette (1982a). The method was received with

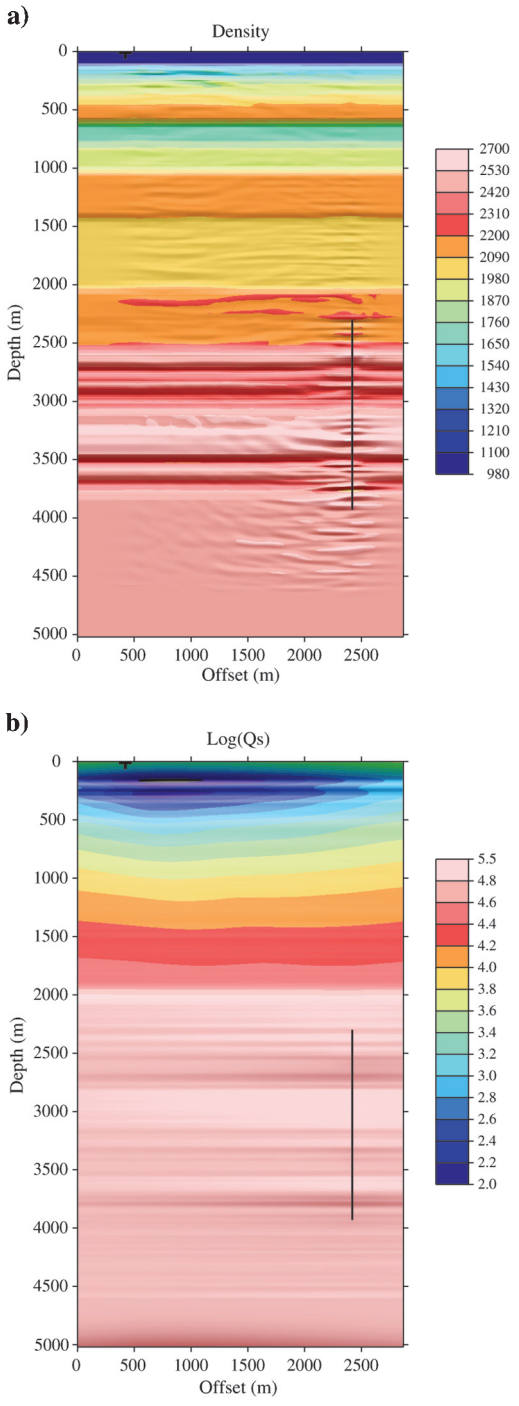

Figure 6. (a) Density model and (b) model obtained from waveform inversion of the data (and the corresponding vertical component data). The density unit is kg/m^3.

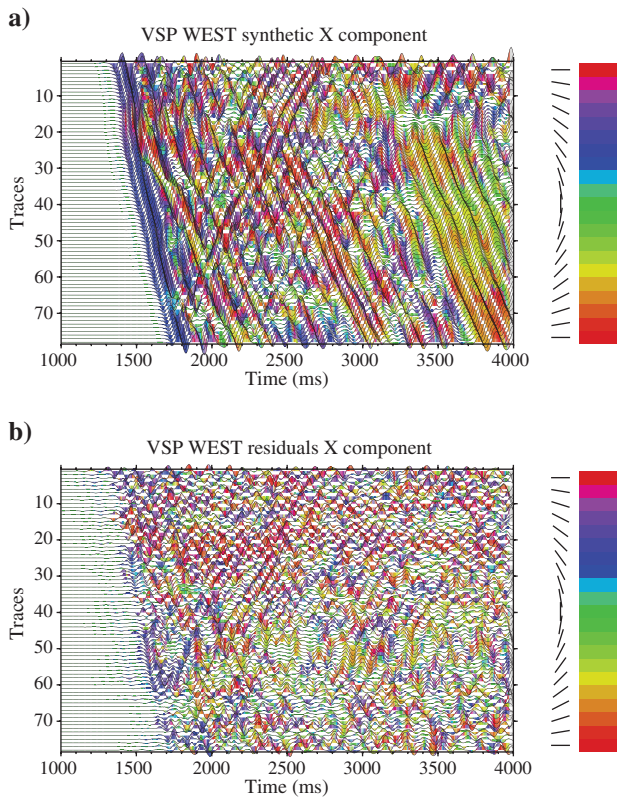

Figure 7. (a) Synthetic data and (b) data residuals after the final iteration of the waveform inversion.

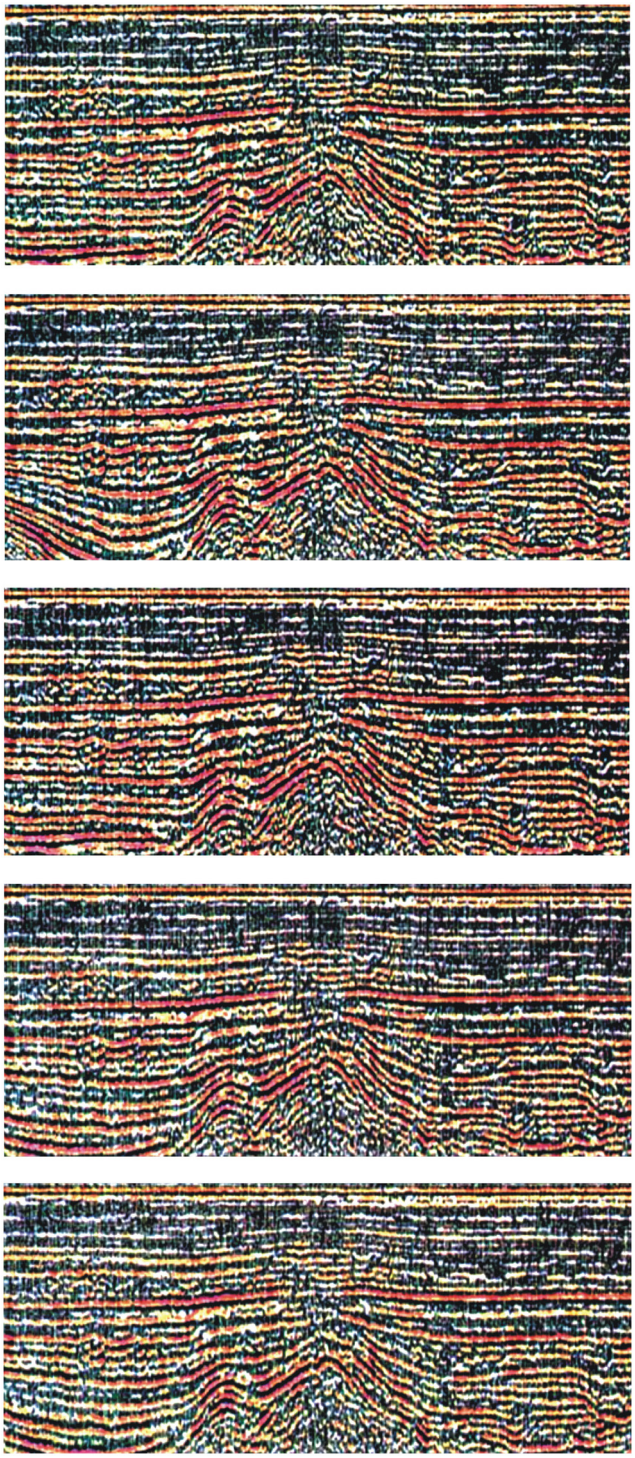

Figure 8. Migrated stacks generated from five posterior samples of a background velocity model for a seismic reflection data set (Koren et al., 1991). Each image shows a vertical cross section of width ca. 14 km and depth 2 km. By displaying a large number of such "posterior images," well-resolved and poorly resolved structures are revealed. Well-resolved structures are preserved in most of the sample models, whereas poorly resolved structures tend to vary between models. Poorly resolved structures can be seen in the lower left part of the image. The figure shows historic photographs taken in the late 1980s from a television set replaying images of the model, hence the poor quality of the figure.

skepticism by many specialists in inverse theory, particularly because its treatment of resolution and uncertainty deviated significantly from the trodden paths of linear inverse theory. Uncertainty and resolution were no longer conceptually separate entities, but together they were represented by frequencies of occurrence of model structure, as seen in the posterior sample. The critics of the late 1990s argued that the method would just produce a meaningless accumulation of models, without giving the user any information about the properties of the solution. Albert maintained that, on the contrary, the method expressed the essence of the inverse problem: a certain feature of the subsurface was "well-resolved" when its frequency of occurrence in the collection of posterior sample models was much larger than in the collection of prior sample models. Its presence was "amplified" in the posterior movie, because its existence was needed to explain the data.

Soon, it became clear that the sampling method had a much larger potential for producing realistic solutions to inverse problems than believed in the first place. Together with Miguel Bosch, Albert developed one of the first methods for rock property inversion, based on MCMC methods (Bosch, 1998). The method was novel, because it formulated the inverse problem as a quest, not for physical parameters of the rocks, but directly for their lithological classification. The property parameters were related to the data through a combination of empirical data (empirical distributions of physical parameters for each rock type) and physical relations (connecting rock physical parameters with the geophysical data). This type of problem was unsolvable with traditional methods, but with sampling methods it became possible. See also Khan et al. (2007).

Later in the 2000s, a new and exciting possibility of exploiting sampling methods for producing realistic solutions appeared. From his very first writings on seismic waveform inversion, Albert Tarantola had the vision that realistic prior information for inversions could be learned from a large collection of "training images" of the subsurface. In the early years of waveform inversion this was, however, not realizable because of limited computer resources, and because the necessary learning algorithms were not yet developed.

It was not until 2003, when Albert met the founder of modern Monte Carlo-based geostatistics, Andre Journel from Stanford University, that events unfolded. They were both invited by Jean-Laurent Mallet (the inventor of Gocad) to a thesis defense in Nancy. They made the two-way travel by train from Paris and Nancy and had ample time to discuss ideas.

Journel defended what he called the "algorithm-driven" approach to define probability models, and this viewpoint was closely related to Albert's nonparametric vision of defining prior probability distributions only from their samples. The result of these discussions was that Andre Journel invited Albert Tarantola to spend three months at Stanford to teach a course on probability and inverse problems. In this period, Albert and Andre took new steps toward integrating modern geostatistical methods with geophysical inversion.

Albert's idea was to exploit the sampling methods of modern geostatistics to generate realistic earth models as prior information in a sampling-based inversion. The Stanford group had developed methods based on sequential simulation (Strebelle, 2002) which allowed them to efficiently generate geologically realistic models, constrained on statistical properties derived from a series of "training images" of relevant geological examples created by, e.g., experienced geologists. This scenario was exactly what Albert had looked for.

Consider a very simple inverse problem where the data are borehole measurements, say porosities, in 50 wells, as shown in Figure 9a. The unknown model is a porosity map over the entire area. It must be consistent with the well data, and at the same time be consistent with the prior: a stochastic process obtained from statistical analysis of training images of possible porosity maps. In this simple example, the prior is a 2D Gaussian process, but in general the prior can be more complex. Figure 9b, 9c, and 9d show three realizations of the posterior probability density, and by displaying such realizations the analyst gains an insight into acceptable realistic structures.

To Albert, the analysis of posterior realizations was essential, because it allowed discovery of well-resolved structure that would otherwise remain hidden if only statistics (mean models, covariances, higher-order moments, etc.) were displayed. The common practice of computing least-squares solutions (posterior mean models or "best-fitting" models) was particularly criticized by Albert because it produces models that look unrealistic. Figure 10 shows the posterior mean model for our simple inverse problem. The difference between the structure of this model and typical sample models like those of Figure 9b, 9c, and 9d is apparent. It was this observation that eventually led Albert to reject the least-squares method in favor of sampling strategies.

During the following years, it became clear that the integration of geophysical and geostatistical data was not easily carried through for large-scale, nonlinear inverse problems. Considerable progress in seismic waveform inversion methods had taken place (see, e.g., Pratt, 1999), but simultaneously constraining subsurface models by full seismic waveforms and complex geostatistical models seemed to be an overwhelming computational task. Albert's original vision had run into difficulties, but he insisted on the feasibility of the approach. Many times before, Albert's persistence had paid off, and in this case he preserved his optimism. His hope was that it would be possible to find a semi-deterministic way of finding solutions that (1) fit the data acceptably, and (2) have a high prior probability, as defined by the stochastic process derived from the training images. Recently, there are signs that important progress is being made toward a solution to this problem (Suzuki and Caers, 2008).

EPILOGUE

Albert Tarantola dedicated much of his life and work to geoscience, but as a scientist his vision was more far-reaching and manifold. In his early years, general relativity taught him to strive toward consistency and invariance, and later in his life inverse problem theory stimulated his fascination for probability as an epistemological instrument. These two themes became important guiding lights for the rest of his career where he devoted much time and enthusiasm to a number of basic problems, including new formulations of fluid dynamics, elasticity theory, and global positioning.

One of his major endeavors was his work on invariant formulations of the laws of physics, that is, formulations independent of parameterizations or coordinate systems. The aim of this ambitious project was to extend the relativistic concept of invariance to mathematical physics in general, and in 2006 he wrote the book *Elements for Physics* presenting the theory. The main thesis of the book was that, although most current physical theories had been reformulated in an invariant language, this was not the case for theories like linear elasticity and the Fourier theory of heat conduction. For example, in the theory of elasticity, Hooke's law is still expressed through a particular choice of parameters/quantities: stress, stiffness, and strain. But a formulation using instead the exponential of the strain, or the inverse of the stiffness, would lead to equations of a fundamentally different form. Albert criticized the inconsistency of such theories. He showed that, under certain circumstances, they lead to unphysical results, and he proposed a way to recast them in a new and invariant formulation.

In a third and unfinished book — *Mapping of Probabilities* — Albert attempted to build a new theory of measurement, free of contradictions. This required a clarification of some of the basic concepts of probability theory (especially the problems related to the Borel paradox), but he also wanted to start afresh and rethink the theory of inference, going back to set theory as the starting

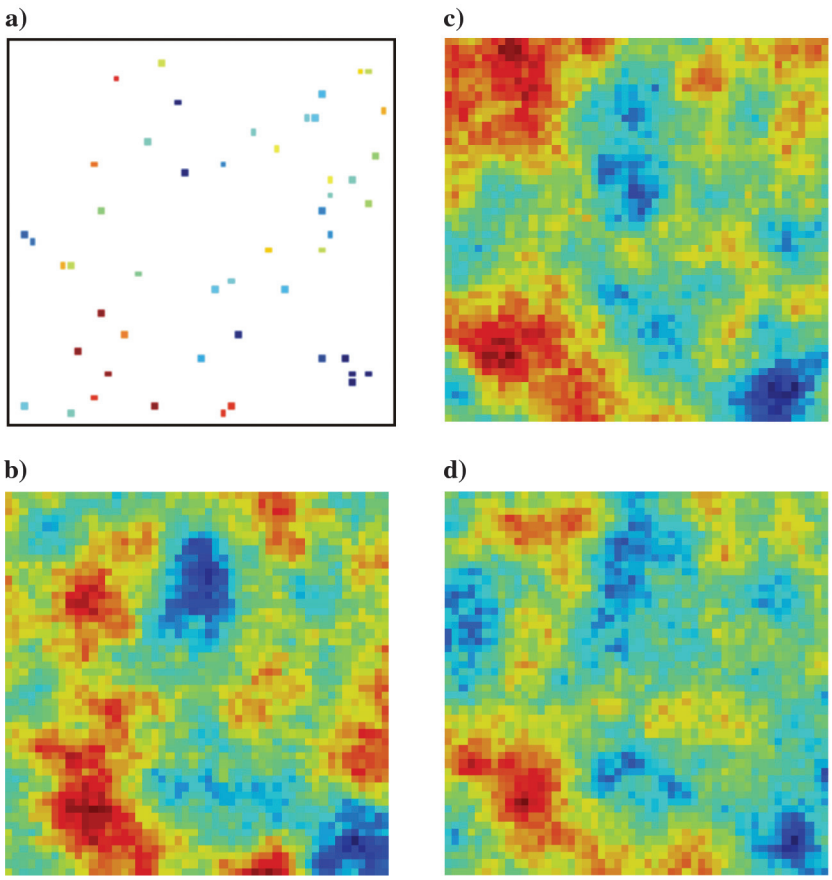

Figure 9. (a) Data, measured at well locations. The data values are shown in color (arbitrary scale with low values in blue and high values in red). (b-d): Earth models, modeled as realizations of 2D Gaussian stochastic process, expressing the a priori information. The realizations were generated by sequential simulation, and were constrained by the well data shown in (a).

point. With this book, he was preparing a new revision of probabilistic inference. Like his old opponents, the frequentists, Albert was critical of the Bayesian paradigm, but for completely different reasons.

These two little-known books show Albert as an independent thinker. Albert did not follow the mainstream of science. He followed his own ideals, although it sometimes led to controversial conclusions. The ideas laid out in *Elements for Physics* were created without much interaction with the physics community. This added to the originality of the work, but it also made it hard for him to publish it.

Contrary to mainstream thinking, Albert believed that scientists should publish only when their ideas were fully matured and worked through. They should strive for scientific quality measured by pure academic standards and not aim at maximizing the number of publications, citations, or patents. Late in his career, he carefully followed these principles. He rejected work that he thought would only improve slightly on existing ideas, and work that avoided discussions about fundamental principles.

Albert possessed a unique combination of exceptional skills and remarkable mental energy. Those who have spent time working with Albert in his favorite Parisian café, Café Beaubourg, know the enthusiasm and joy he radiated during these sessions. After many hours of concentrated work with his companions, he would go home and sum up the discussions on dozens of densely written and carefully typeset pages, supplemented with new material and new insights. He was a veritable powerhouse with an unusual work ethic, even for a person at his scientific level.

Albert was persistent, and in many cases it was worthwhile. In much of his work, he was many years ahead of computational possibilities. He supported development of Monte Carlo methods long before they were accepted as useful methods, and today we see large-scale waveform inversion of seismic data emerging as a real possibility.

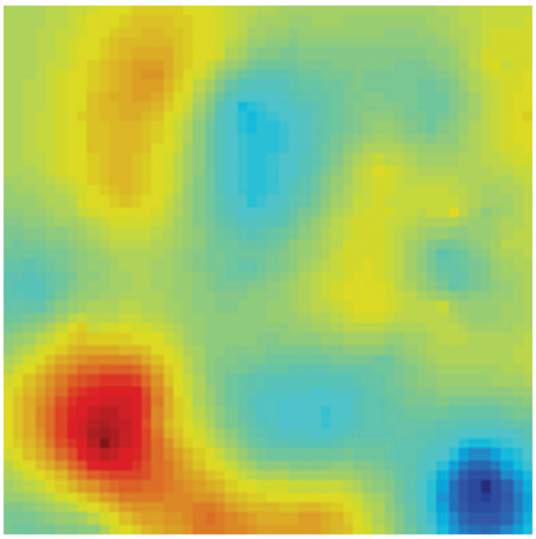

Figure 10. The least-squares solution to the problem of interpolating between the well data shown in Figure 9a. The fact that this solution is a very unlikely outcome of the stochastic process, whose realizations are illustrated in Figure 9b, 9c, and 9d, led Albert to reject the least-squares technique as a satisfactory method.

With Albert's death, the scientific community has lost a unique scientist and person. Albert died much too early, still actively pursuing his scientific goals. He will be missed by family, friends, and colleagues.

ACKNOWLEDGMENTS

The author would like to thank Sven Treitel, John Scales, Roel Snieder, Zvi Koren, and an anonymous reviewer for their very insightful and personal comments to the manuscript. The author was supported by The Danish Council for Independent Research — Natural Sciences (FNU), Grant 272-08-0568.

REFERENCES

Backus, G., and F. Gilbert, 1968, The resolving power of gross Earth data: Geophysical Journal of the Royal Astronomical Society, **16**, 169–205.
Backus, G., and F. Gilbert, 1970, Uniqueness in the inversion of inaccurate gross Earth data: Philosophical Transactions of the Royal Society of London, **266**, 123–192, doi: 10.1098/rsta.1970.0005.
Backus, G. E., 1988, Comparing hard and soft prior bounds in geophysical inverse problems: Geophysical Journal International, **94**, 249–261.
Barnes, C., M. Charara, and A. Tarantola, 1998, Monte Carlo inversion of arrival times for multiple phases in OSVP data: 68th Annual International Meeting, SEG, Expanded Abstracts, 1871–1874.
Bernardo, J. M., and F. M. Smith, 1994, Bayesian Theory: Wiley.
Bosch, M., 1998, Tomographie lithologique: methode et applications (in French): Ph.D. thesis, Institut de Physique du Globe, Paris.
Charara, M., C. Barnes, and A. Tarantola, 1996, The state of affairs in inversion of seismic data: An OVSP example: 66th Annual International Meeting, SEG, Expanded Abstracts, 1999-2002.
Charara, M., C. Barnes, and A. Tarantola, 2000, Full waveform inversion of seismic data for a visco-elastic medium: in Methods and Applications of Inversion, Lecture Notes in Earth Sciences, **92**, Springer-Verlag, New York.
Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481, doi: 10.1190/1.1440185.
Crase, E., C. Wideman, M. Noble, and A. Tarantola, 1992, Nonlinear elastic waveform inversion of land seismic reflection data: Journal of Geophysical Research, **97**, no. B4, 1992, 4685–4703, doi: 10.1029/90JB00832.
Franklin, J. N., 1970, Well posed stochastic extensions of ill posed linear problems: Journal of Mathematical Analysis and Applications, **31**, 682–716, doi: 10.1016/0022-247X(70)90017-X.
Jackson, D. D., and M. Matsu'ura, 1985, A Bayesian approach to nonlinear inversion: Journal of Geophysical Research, **90**, B1, 581–591, doi: 10.1029/JB090iB01p00581.
Jaynes, E. T., 1968, Prior probabilities: IEEE Transactions on systems, science, and cybernetics, SSC-**4**, no. 3, 227–241.
Jeffreys, H., 1939, Theory of probability: Clarendon Press, Oxford. Reprinted in 1961 by Oxford University Press.
Khan, A., J. A. D. Connolly, J. Maclennan, and K. Mosegaard, 2007, Joint inversion of seismic and gravity data for lunar composition and thermal state: Geophysical Journal International.
Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, **96**, no. B12, doi: 10.1029/91JB02278.
Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migration, *in* J. B. Bednar, R. Rednar, E. Robinson, and A. Weglein, eds., Conference on Inverse Scattering: Theory and Application, SIAM.
Lailly, P., 2010, Non-linear inversion: who was first? Paper presented at the Albert Tarantola Homage Workshop.
Levenberg, K., 1944, A method for the solution of certain nonlinear problems in least-squares: Quarterly of Applied Mathematics, **2**, 164–168.
Marquardt, D. W., 1970, Generalized inverses, ridge regression, biased linear estimation and non-linear estimation: Technometrics, **12**, 591–612, doi: 10.2307/1267205.
Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, no. B7, 12,431–12,447, doi: 10.1029/94JB03097.
Mosegaard, K., and A. Tarantola, 2002, Probabilistic Approach to Inverse Problems: Chapter for the International Handbook of Earthquake and Engineering Seismology: Academic Press, 237–265. 0-12-440652-1. Published for the International Association of Seismology and Physics of the Earth Interior (IASPEI).
Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: theory and verification in a physical scale model: Geophysics, **64**, 888–901, doi: 10.1190/1.1444597.

Rothman, D. H., 1986, Automatic estimation of large residual statics corrections: Geophysics, **51**, 332–346, doi: 10.1190/1.1442092.
Shannon, C. E., 1948, A mathematical theory of communication: Bell System Technical Journal, **27**, 379–423.
Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics: Mathematical Geology, **34**, 1, 1–21, doi: 10.1023/A:1014009426274.
Suzuki, S., and J. Caers, 2008, A distance-based prior model parameterization for constraining solutions of spatial inverse problems: Mathematical Geosciences, **40**, 4, 445–469, doi: 10.1007/s11004-008-9154-8.
Talagrand, O., and P. Courtier1987 Variational assimilation of meterological observations with the adjoint vorticity equation: Theory: Quarterly Journal of the Royal Meteorological Society
Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.
Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903, doi: 10.1190/1.1442046.
Tarantola, A., 1987, Inverse problem theory; methods for data fitting and model parameter estimation: Elsevier.
Tarantola, A., 1988, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation: Pure and Applied Geophysics, **128**, 1-2, 365–399, doi: 10.1007/BF01772605.
Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM, 163–169.
Tarantola, A., 2006, Elements for physics — Quantities, qualities, and intrinsic theories: Springer-Verlag Berlin Heidelberg.
Tarantola, A., and B. Valette, 1982a, Inverse Problems = Quest for Information: Journal of Geophysics, **50**, 159–170.
Tarantola, A., and B. Valette, 1982b, Generalized nonlinear inverse problems solved using the least-squares criterion: Reviews of Geophysics and Space Physics, **20**, 2, 219–232, doi: 10.1029/RG020i002p00219.
Tikhonov, A. N., 1963, Resolution of ill-posed problems and the regularization method (in Russian): Doklady Akademii Nauk SSSR, **151**, 501–504.
Vestergaard, P. D., and K. Mosegaard, 1991, Inversion of post-stack seismic data using simulated annealing: Geophysical Prospecting, **39**, 613–624, doi: 10.1111/gpr.1991.39.issue-5.
Zuk, T., and S. et Carpendale, 2006, Theoretical analysis of uncertainty visualization: Proceedings of SPIE, SPIE-6060, 66–79.